The Linear Capacity of Single-Server Individually-Private Information Retrieval with Side Information

Anoosheh Heidarzadeh and Alex Sprintson

Abstract— This paper considers the problem of single-server Individually-Private Information Retrieval with side information (IPIR). In this problem, there is a remote server that stores a dataset of K messages, and there is a user that initially knows M of these messages, and wants to retrieve D other messages belonging to the dataset. The goal of the user is to retrieve the D desired messages by downloading the minimum amount of information from the server while revealing no information about whether an individual message is one of the D desired messages. In this work, we focus on linear IPIR schemes, i.e., the IPIR schemes in which the user downloads only linear combinations of the original messages from the server. We prove a converse bound on the download rate of any linear IPIR scheme for all K, D, M, and show the achievability of this bound for all K, D, M satisfying a certain divisibility condition. Our results characterize the linear capacity of IPIR, which is defined as the maximum achievable download rate over all linear IPIR schemes, for a wide range of values of K, D, M.

I. INTRODUCTION

In this work, we consider the problem of *single-server Individually-Private Information Retrieval with side information*, which we refer to as *IPIR* for short. In this problem, there is a set of K messages stored on a remote server, and there is a user that has M (out of K) messages as side information, and wants to retrieve D other messages. The objective is to design a retrieval scheme in which the user downloads the minimum possible amount of information from the server while revealing no information about the identity of every individual message required by the user.

The IPIR problem, which was originally introduced in [1] and later studied in [2], is related to several work in the Private Information Retrieval (PIR) literature. In particular, the IPIR problem is a variant of the problem of multi-message PIR with side information (MPIR-SI) which is a generalization of the multi-message PIR problem [3], [4]. In the MPIR-SI problem, a user wishes to privately retrieve multiple messages, with the help of a prior side information, from a single (or multiple) remote server(s) storing (identical copies or coded versions of) a set of messages.

The MPIR-SI problem has been studied under three different information-theoretic privacy guarantees: *full privacy*, *joint privacy*, and *individual privacy*. In the case of full privacy, both the identities of the messages required by the user and the identities of the user's side information messages must be kept private from the server(s). In contrast, when joint or individual privacy is required, only the identities of the messages required by the user must be protected,

The authors are with the Department of Electrical and Computer Engineering, Texas A&M University, College Station, TX 77843 USA (E-mail: {anoosheh, spalex}@tamu.edu).

and it is not required to protect the identities of the user's side information messages. In the case of joint privacy, the server(s) must not learn which subset of messages was required by the user, whereas in the case of individual privacy, the server(s) must not learn whether an individual message was one of the messages required by the user.

The joint privacy guarantee finds application in scenarios in which the correlation between the identities of the required messages must be protected (see, e.g., [5], [6]), whereas the individual privacy guarantee is of practical importance in scenarios in which there is no need to protect the correlation between the identities of the required messages (see, e.g., [7], [8]). Note that the joint and individual privacy requirements are equivalent in the classical PIR problem and the problem of PIR with side information where the user wants to privately retrieve only one message [9]–[26].

Several variants of the MPIR-SI problem were previously studied in the literature. The single-server setting of MPIR-SI with full privacy was studied in [27], and the multi-server setting of this problem was studied in [28]. In addition, the single-server setting of MPIR-SI with joint privacy, which we refer to as *JPIR* for short, was studied in [27], [29], and the single-server setting of MPIR-SI with individual privacy, which is the IPIR problem, was considered in [1], [2].

In [1], we proposed an IPIR scheme for all K, D, M, and showed that this scheme achieves a download rate higher than that of the JPIR schemes of [27] and [29]. The optimality of the scheme of [1] was also shown for $D = 2, M = 1$, but it was left open in general. Recently, in [2], we showed that the scheme of [1] is not always optimal, and proposed an optimal IPIR scheme for $D = 2, M = 2$ that achieves a download rate higher than that of the scheme of [1]. Notwithstanding, the fundamental limits of the IPIR problem have remained unknown for all other values of D, M.

In this work, we focus on linear IPIR schemes, i.e., the user downloads only linear combinations of the original messages from the server. We prove a converse bound on the download rate of any linear IPIR scheme for all K, D, M, and show the achievability of this bound for all K, D, M satisfying a certain divisibility condition. Our results characterize the optimal download rate of linear IPIR and show the suboptimality of the IPIR scheme of [1], for a wide range of values of K, D, M. Our converse proof technique relies on a mix of combinatorial, algebraic, and information-theoretic arguments that are tailored to the single-server setting, linear retrieval schemes, and the individual privacy guarantee. In addition, our achievability scheme is based on randomized partitioning and maximum distance separable (MDS) codes.

II. PROBLEM SETUP

Throughout, we denote random variables and their realizations by bold-face symbols and regular symbols, respectively. For any integer $i \geq 1$, we denote $\{1, \ldots, i\}$ by $[i]$. Also, we denote the binomial coefficient $\binom{n}{k}$ by $C_{n,k}$.

Let $\mathbb{F}_q$ be a finite field of order q, and let $\mathbb{F}_q^n$ be the n-dimensional vector space over $\mathbb{F}_q$. Let K, D, M be arbitrary integers such that $D \geq 2$, $M \geq 1$, and $K \geq D + M$.

Consider a server that stores K messages $X_1, \ldots, X_K$, where $X_i = [X_{i,1}, \ldots, X_{i,n}] \in \mathbb{F}_q^n$ for $i \in [K]$. Note that $X_{i,j} \in \mathbb{F}_q$ for all $i \in [K]$ and all $j \in [n]$. We refer to $X_{i,1}, \ldots, X_{i,n}$ as the *symbols* of the message X_i. For simplifying the notation, we denote by $X_{\mathcal{I}}$ the set of messages $\{X_i : i \in \mathcal{I}\}$ for every $\mathcal{I} \subseteq [K]$.

Let $\mathbb{W}$ be the set of all D-subsets of $[K]$, and let $\$$ be the set of all M-subsets of $[K]$. Consider a user who initially knows the M messages X_S for a given $S \in \$$, and wishes to retrieve the D messages X_W for a given $W \in \mathbb{W}$. To avoid degenerate cases, we assume that $W \cap S = \emptyset$. We refer to X_W as the *demand*, X_S as the *side information*, W as the *index set of the demand*, S as the *index set of the side information*, D as the *size of the demand*, and M as the *size of the side information*.

In this work, we make the following assumptions:

1) $X_1, \ldots, X_K$ are independent and uniformly distributed over $\mathbb{F}_q^n$. Thus, $H(X_{\mathcal{I}}) = |\mathcal{I}|B$ for all $\mathcal{I} \subseteq [K]$, where $B := n \log_2 q$ is the entropy of a message (in bits).
2) $(\mathbf{W}, \mathbf{S})$ and $X_1, \ldots, X_K$ are independent.
3) The distribution of $\mathbf{S}$ is uniform over all $S \in \$$, and the conditional distribution of $\mathbf{W}$ given $\mathbf{S} = S$ is uniform over all $W \in \mathbb{W}$ such that $W \cap S = \emptyset$. Thus, $\mathbb{P}(\mathbf{W} = W, \mathbf{S} = S) = 1/C_{K,M} \times 1/C_{K-M,D}$ for all $(W, S) \in \mathbb{W} \times \$$ such that $W \cap S = \emptyset$. Moreover, $\mathbb{P}(i \in \mathbf{W}) = \sum_{W \in \mathbb{W}: i \in W} \mathbb{P}(\mathbf{W} = W) = \sum_{W \in \mathbb{W}: i \in W} \sum_{S \in \$: W \cap S = \emptyset} \mathbb{P}(\mathbf{W} = W, \mathbf{S} = S) = C_{K-1,D-1} \times C_{K-D,M} \times 1/C_{K,M} \times 1/C_{K-M,D} = D/K$ for all $i \in [K]$.
4) The demand's size D, the side information's size M, and the distribution of $(\mathbf{W}, \mathbf{S})$ are initially known by the server, whereas the realization (W, S) is initially unknown to the server.

Given (W, S), the user generates a query $Q^{[W,S]}$, and sends it to the server. The query $Q^{[W,S]}$ is a deterministic or stochastic function of (W, S), independent of $X_1, \ldots, X_K$. Given the query, every message index must be equally likely to belong to the demand's index set. That is, for every $i \in [K]$, we must have $\mathbb{P}(i \in \mathbf{W} | \mathbf{Q}^{[\mathbf{W},\mathbf{S}]} = Q^{[W,S]}) = \mathbb{P}(i \in \mathbf{W}) = D/K$. We refer to this condition as the *individual privacy condition*.

Upon receiving $Q^{[W,S]}$, the server generates an answer $A^{[W,S]}$, and sends it back to the user. The answer $A^{[W,S]}$ is a deterministic function of $Q^{[W,S]}$ and $X_1, \ldots, X_K$. That is, $H(\mathbf{A}^{[\mathbf{W},\mathbf{S}]} | \mathbf{Q}^{[\mathbf{W},\mathbf{S}]}, X_1, \ldots, X_K) = 0$. The user must be able to recover the demand X_W given the answer $A^{[W,S]}$, the query $Q^{[W,S]}$, the side information X_S, and the realization (W, S). That is, $H(X_W | \mathbf{A}^{[\mathbf{W},\mathbf{S}]}, \mathbf{Q}^{[\mathbf{W},\mathbf{S}]}, X_S) = 0$. We refer to this condition as the *recoverability condition*.

The problem is to design a protocol for generating a query $Q^{[W,S]}$ and the corresponding answer $A^{[W,S]}$ for any given (W, S) such that both the individual privacy and recoverability conditions are satisfied. We refer to this problem as *single-server Individually-Private Information Retrieval with side information*, or *IPIR* for short.

In this work, we focus on *(scalar-) linear* protocols, i.e., any protocol in which the server's answer $A^{[W,S]}$ consists only of linear combinations of the messages $X_1, \ldots, X_K$ (with combination coefficients from $\mathbb{F}_q$). Note that for any linear protocol, the user's query $Q^{[W,S]}$ can be fully specified by the corresponding combination coefficient vectors. Without loss of generality, we assume that these combination coefficient vectors are linearly independent (over $\mathbb{F}_q$).

We define the *rate* of a linear IPIR protocol as the ratio of the amount of information required by the user to the amount of information downloaded from the server, i.e., $H(X_W)/H(\mathbf{A}^{[\mathbf{W},\mathbf{S}]})$, and define the *linear capacity* of IPIR as the supremum of rates over all linear IPIR protocols. By the assumptions (1) and (2), it can be shown that (i) $H(X_W) = DB$, and (ii) $H(\mathbf{A}^{[\mathbf{W},\mathbf{S}]}) = H(\mathbf{L}^{[\mathbf{W},\mathbf{S}]}) + \mathbb{E}[\mathbf{L}^{[\mathbf{W},\mathbf{S}]}]B$, where $\mathbf{L}^{[\mathbf{W},\mathbf{S}]}$ represents the number of linear combinations that constitute the answer $\mathbf{A}^{[\mathbf{W},\mathbf{S}]}$. In particular, if $\mathbf{L}^{[\mathbf{W},\mathbf{S}]}$ is a constant random variable taking only the value L for some integer $L \geq 1$ (i.e., all realizations of $\mathbf{A}^{[\mathbf{W},\mathbf{S}]}$ consist of L linear combinations), then $H(\mathbf{A}^{[\mathbf{W},\mathbf{S}]}) = LB$, and the rate of the protocol is D/L. It should be noted that (i) and (ii) do not rely on the assumptions (3) and (4).

In this work, our goal is to characterize the linear capacity of IPIR in terms of the parameters K, D, M.

III. MAIN RESULTS

This section summarizes our main results. Theorem 1 provides an upper bound on the linear capacity of IPIR for all K, D, M, and Theorem 2 provides a matching lower bound on the linear capacity of IPIR for all K, D, M such that M is an integer multiple of D. The proof of converse and achievability are presented in Sections IV and V, respectively.

Theorem 1. *For IPIR with K messages, demand's size D, and side information's size M, the linear capacity is upper bounded by $(D + M)/K$.*

We present a novel technique to prove the converse when the number of symbols of a message (n) grows unbounded. It should be noted that the converse bound holds for any $n \geq 1$. This is because the linear capacity for any finite n cannot exceed the linear capacity as n tends to infinity. (Any linear protocol for $n = n_0$ can serve as a linear protocol—achieving the same rate—for $n = k n_0$ for any arbitrary integer $k \geq 1$.) Our proof technique is based on a mix of combinatorial, algebraic, and information-theoretic arguments which rely on the individual-privacy and recoverability conditions. The main idea of the proof is to show that given the answer and the query of any linear IPIR protocol, there exists a collection of coded symbols (each coded symbol being a linear combination of the symbols of a message with combination coefficients from $\mathbb{F}_q$) of size at most $(M/(D + M))Kn$ from which all Kn message symbols can be recovered.

To prove this, we use a standard random linear network coding argument and constructively identify the number of required coded symbols for each message. This result implies that the amount of information downloaded from the server in any linear IPIR protocol is lower bounded by $KB - (M/(D+M))KB = (D/(D+M))KB$ bits, where $B = n \log_2 q$ is the amount of information in a message (in bits). Since the amount of information required by the user is DB bits, then the rate of any linear IPIR protocol is upper bounded by $DB/((D/(D+M))KB) = (D+M)/K$.

Theorem 2. *For IPIR with K messages, demand's size D, and side information's size M, the linear capacity is lower bounded by $(D+M)/K$ when $(D+M) \mid K$, or more generally, when $((D+M)/\gcd(D,M)) \mid K$.*

To prove the achievability result, we propose a capacity-achieving linear IPIR protocol, which we call *Group-and-Code*, for all K, D, M such that $D/R + M/R$ is an integer divisor of K, where $R := \gcd(D,M)$. This protocol is applicable for any $n \geq 1$ and any $q \geq 2$ such that there exists a $[D/R + M/R, D/R]$ MDS code over $\mathbb{F}_q$ (e.g., any $q \geq 2$ or any $q \geq D/R + M/R$ when $D/R = 1$ or $D/R > 1$, respectively.) The key idea of the proposed protocol is to carefully divide the messages into $K/(D/R + M/R)$ disjoint groups of size $D/R + M/R$, and query D/R MDS-coded combinations of the messages in each group.

Remark 1. The results of Theorems 1 and 2 show that the linear capacity of IPIR is given by $(D+M)/K$ for all K, D, M such that $D/R + M/R$ divides K. The highest rate previously shown to be achievable for IPIR [1] was $D/(K - M\lfloor K/(D+M)\rfloor)$ for all K, D, M such that $(K-D)/(D+M) \leq \lfloor K/(D+M)\rfloor$, and $1/\lceil K/(D+M)\rceil$ for all K, D, M such that $(K-D)/(D+M) > \lfloor K/(D+M)\rfloor$. This achievable rate reduces to $(D+M)/K$ when $(D+M) \mid K$, yet it is strictly lower than $(D+M)/K$ when $(D+M) \nmid K$. In contrast, the proposed scheme in this work achieves the rate $(D+M)/K$ for a wider range of values of K, D, M, particularly when $(D+M) \nmid K$ but $(D/R + M/R) \mid K$. Not only does this show the sub-optimality of the IPIR scheme of [1] in general, but it also shows the significance of the results in this work. We conjecture that the linear capacity of IPIR is given by $D/\lceil DK/(D+M)\rceil$ for all K, D, M, which reduces to $(D+M)/K$ when $(D/R + M/R) \mid K$.

Remark 2. The (general) capacity of IPIR, which is defined as the maximum achievable rate over all linear and non-linear IPIR protocols, was previously characterized for the two cases of $D = 2, M = 1$ and $D = 2, M = 2$ in [1] and [2], respectively. In particular, the capacity was shown to be $2/\lceil 2K/3\rceil$ and $2/\lceil K/2\rceil$ for $D = 2, M = 1$ and $D = 2, M = 2$, respectively. In addition, it was shown that in both of these cases the capacity can be achieved by a linear protocol. This shows that the linear capacity of IPIR and the general capacity of IPIR for these cases are the same. We conjecture that the general capacity of IPIR does not exceed the linear capacity of IPIR for any K, D, M, and it is given by $D/\lceil DK/(D+M)\rceil$ for all K, D, M.

Remark 3. The linear capacity of IPIR, which is given by $(D+M)/K$ for all K, D, M such that $(D/R + M/R) \mid K$, may approach zero as K tends to infinity, e.g., when D and M are constant with respect to K. However, the linear capacity of IPIR does not always approach zero as K grows unbounded. For instance, when $D = \alpha K$ and $M = \beta K$ for arbitrary constants $0 < \alpha, \beta < 1$ such that $\alpha + \beta \leq 1$ (i.e., D and M grow linearly in K), as K grows unbounded, the linear capacity of IPIR approaches the nonzero constant $\alpha + \beta$, which can be as large as 1.

IV. PROOF OF THEOREM 1

For simplifying the notation, we denote $\mathbf{Q}^{[\mathbf{W},\mathbf{S}]}$ and $\mathbf{A}^{[\mathbf{W},\mathbf{S}]}$ by $\mathbf{Q}$ and $\mathbf{A}$, respectively. We need to show that $H(\mathbf{A}) \geq (D/(D+M))KB$, where $B = n \log_2 q$ is the entropy of a message (in bits). Recall that each message X_i consists of n independent and uniformly distributed symbols $\{X_{i,j}\}_{j \in [n]}$ over $\mathbb{F}_q$. Let $N := q^n - 1$, and let $c_1, \ldots, c_N$ be the set of all nonzero vectors in $\mathbb{F}_q^n$. For each $m \in [N]$, let $c_m = [c_{m,1}, \ldots, c_{m,n}]$. Note that $c_{m,j} \in \mathbb{F}_q$ for all $m \in [N]$ and for all $j \in [n]$. We refer to $Y_{i,m} := \sum_{j \in [n]} c_{m,j} X_{i,j}$ as a *coded symbol* of the message X_i. To prove that $H(\mathbf{A}) \geq (D/(D+M))KB$, it suffices to show that for any linear IPIR protocol, given the query and the answer, there exist $R \leq (M/(D+M))Kn$ coded symbols $\{Y_{i,m}\}_{\mathcal{T}} := \{Y_{i,m} : (i,m) \in \mathcal{T}\}$ for some R-subset $\mathcal{T}$ of $[K] \times [N]$ (depending on the realization of the query and the answer) given which all Kn message symbols $\{X_{i,j}\} := \{X_{i,j} : i \in [K], j \in [n]\}$ can be recovered, i.e., $H(\{\mathbf{X}_{i,j}\}|\mathbf{A}, \mathbf{Q}, \{\mathbf{Y}_{i,m}\}_{\mathcal{T}}) = 0$. This is because

$$
\begin{aligned}
H(\mathbf{A}) &\geq H(\mathbf{A}|\mathbf{Q}, \{\mathbf{Y}_{i,m}\}_{\mathcal{T}}) \\
&\overset{(a)}{=} H(\mathbf{A}|\mathbf{Q}, \{\mathbf{Y}_{i,m}\}_{\mathcal{T}}) + H(\{\mathbf{X}_{i,j}\}|\mathbf{A}, \mathbf{Q}, \{\mathbf{Y}_{i,m}\}_{\mathcal{T}}) \\
&\overset{(b)}{=} H(\{\mathbf{X}_{i,j}\}|\mathbf{Q}, \{\mathbf{Y}_{i,m}\}_{\mathcal{T}}) + H(\mathbf{A}|\mathbf{Q}, \{\mathbf{X}_{i,j}\}) \\
&\overset{(c)}{=} H(\{\mathbf{X}_{i,j}\}|\mathbf{Q}, \{\mathbf{Y}_{i,m}\}_{\mathcal{T}}) \\
&\overset{(d)}{=} H(\{\mathbf{X}_{i,j}\}|\{\mathbf{Y}_{i,m}\}_{\mathcal{T}}) \\
&\overset{(e)}{\geq} H(\{\mathbf{X}_{i,j}\}) - H(\{\mathbf{Y}_{i,m}\}_{\mathcal{T}}) \\
&\overset{(f)}{\geq} (Kn - R)B/n \\
&\overset{(g)}{\geq} (D/(D+M))KB,
\end{aligned}
$$

where (a) holds because $H(\{\mathbf{X}_{i,j}\}|\mathbf{A}, \mathbf{Q}, \{\mathbf{Y}_{i,m}\}_{\mathcal{T}}) = 0$ by assumption; (b) follows from the chain rule of entropy; (c) holds because $H(\mathbf{A}|\mathbf{Q}, \{\mathbf{X}_{i,j}\}) = 0$; (d) follows because $\mathbf{Q}$ and $\{\mathbf{X}_{i,j}\}$ are independent (by assumption); (e) holds because $H(\{\mathbf{X}_{i,j}\}|\{\mathbf{Y}_{i,m}\}_{\mathcal{T}}) = H(\{\mathbf{X}_{i,j}\}) + H(\{\mathbf{Y}_{i,m}\}_{\mathcal{T}}|\{\mathbf{X}_{i,j}\}) - H(\{\mathbf{Y}_{i,m}\}_{\mathcal{T}})$, and $H(\{\mathbf{Y}_{i,m}\}_{\mathcal{T}}|\{\mathbf{X}_{i,j}\}) = 0$ because $Y_{i,m}$ is a linear combination of $\{X_{i,j} : j \in [n]\}$; (f) holds because $H(\{\mathbf{X}_{i,j}\}) = KB$, noting that $\{\mathbf{X}_{i,j}\}$ are independent and uniformly distributed over $\mathbb{F}_q$, and $H(\{\mathbf{Y}_{i,m}\}_{\mathcal{T}}) \leq |\mathcal{T}|H(\mathbf{Y}_{i,m}) = RB/n$, noting that $\mathbf{Y}_{i,m}$ is a linear combination of $\{X_{i,j} : j \in [n]\}$, and hence uniformly distributed over $\mathbb{F}_q$, i.e., $H(\mathbf{Y}_{i,m}) = \log_2 q = B/n$; and (g) follows because $R \leq (M/(D+M))Kn$ by assumption.

Consider an arbitrary linear IPIR protocol. Fix arbitrary $W \in \mathbb{W}$ and $S \in \mathbb{S}$ such that $W \cap S = \emptyset$. Let $Q^{[W,S]}$ and $A^{[W,S]}$ be a query and its corresponding answer generated by the protocol, respectively. For ease of notation, we denote $Q^{[W,S]}$ and $A^{[W,S]}$ by Q and A, respectively.

For any $(W^*, S^*) \in \mathbb{W} \times \mathbb{S}$ such that $W^* \cap S^* = \emptyset$, we say that the tuple (W^*, S^*) is *feasible* (given Q and A) if X_{W^*} can be recovered given X_{S^*}. Let $(W_1, S_1), \ldots, (W_T, S_T)$ be the set of all feasible tuples given Q and A. By the linearity of the protocol, it follows that for any $l \in [T]$, any $k \in W_l$, and any $j \in [n]$, the message symbol $X_{k,j}$ (i.e., the jth symbol of the message X_k) can be recovered given the M message symbols $\{X_{i,j}\}_{i \in S_l}$. This further implies that for any $k \in W_l$ and any $m \in [N]$, the coded symbol $\sum_{j \in [n]} c_{m,j} X_{k,j}$ can be recovered given the M coded symbols $\{\sum_{j \in [n]} c_{m,j} X_{i,j}\}_{i \in S_l}$.

Recall that we need to show that there exist $R \le (M/(D+M))Kn$ coded symbols given which all Kn message symbols can be recovered. As discussed in Section III, it suffices to prove this claim for sufficiently large n. In the following we present a proof by construction.

For each $l \in [T]$, let p_l be the conditional probability that $\mathbf{W} = W_l$ and $\mathbf{S} = S_l$ given that $\mathbf{Q} = Q$. That is, $p_l := \mathbb{P}(\mathbf{W} = W_l, \mathbf{S} = S_l | \mathbf{Q} = Q)$. Note that $\sum_{l \in [T]} p_l = 1$. For any $i \in [K]$, we define $\mathcal{F}_i := \{l \in [T] : i \in W_l\}$ and $\mathcal{E}_i := \{l \in [T] : i \in S_l\}$. Let $\alpha_i := \sum_{l \in \mathcal{F}_i} p_l$ and $\beta_i := \sum_{l \in \mathcal{E}_i} p_l$ for all $i \in [K]$. Note that $\alpha_i \stackrel{(a)}{=} \mathbb{P}(i \in \mathbf{W} | \mathbf{Q} = Q) \stackrel{(b)}{=} \mathbb{P}(i \in \mathbf{W}) = D/K$, where (a) follows from the law of total probability; and (b) follows from the individual privacy condition. To simplify the notation, we define $\alpha := D/K$. Note also that $\beta_i = \mathbb{P}(i \in \mathbf{S} | \mathbf{Q} = Q)$ (by the law of total probability). Moreover, $\sum_{i \in [K]} \beta_i = M$. This is because $\sum_{i \in [K]} \beta_i = \sum_{i \in [K]} \sum_{l \in \mathcal{E}_i} p_l = \sum_{l \in [T]} M p_l = M$.

Let $\mathcal{I}_0 = \emptyset$, and let $\mathcal{I}_1, \ldots, \mathcal{I}_P$ be a partition of $[K]$ such that for each $k \in [P]$, $\beta_i = \beta_j$ for all $i, j \in \mathcal{I}_k$, and for every $k, h \in [P]$, $\beta_i \neq \beta_j$ for all $i \in \mathcal{I}_k$ and for all $j \in \mathcal{I}_h$. For each $k \in [P]$, let γ_k be such that $\beta_i = \gamma_k$ for all $i \in \mathcal{I}_k$. Without loss of generality, assume that $\gamma_1 > \gamma_2 > \cdots > \gamma_P$. Let $r_0 := 0$, and let $r_k := n/(\alpha + \gamma_k) - \sum_{h=0}^{k-1} r_h$ for all $k \in [P]$.

For sufficiently large n, we show that (i) there exist $\mathcal{N}_{k,l} \subset [N]$, $|\mathcal{N}_{k,l}| = r_k p_l$ for all $k \in [P]$ and for all $l \in [T]$, such that all Kn message symbols $\{X_{i,j} : i \in [K], j \in [n]\}$ can be recovered given the coded symbols $\{Y_{i,m} : i \in [K] \setminus \cup_{h=0}^{k-1} \mathcal{I}_h, m \in \mathcal{N}_{k,\mathcal{E}_i}\}_{k \in [P]}$, and (ii) $|\{Y_{i,m} : i \in [K] \setminus \cup_{h=0}^{k-1} \mathcal{I}_h, m \in \mathcal{N}_{k,\mathcal{E}_i}\}_{k \in [P]}| \le (M/(D+M))Kn$, where $\mathcal{N}_{k,\mathcal{F}_i} := \{\mathcal{N}_{k,l}\}_{l \in \mathcal{F}_i}$ and $\mathcal{N}_{k,\mathcal{E}_i} := \{\mathcal{N}_{k,l}\}_{l \in \mathcal{E}_i}$. From now on, whenever we use the notation $\mathcal{N}_{k,l}$ for any $k \in [P]$ and $l \in [T]$, it is assumed that $\mathcal{N}_{k,l}$ is a subset of $[N]$ of size $r_k p_l$, noting that $r_k p_l$ is an integer, for sufficiently large n.

We say that a collection $\{\mathcal{N}_{1,l}\}_{l \in [T]}$ is *good* if all message symbols $\{X_{i,j} : i \in \mathcal{I}_1, j \in [n]\}$ can be recovered given the coded symbols $\{Y_{i,m} : i \in [K], m \in \mathcal{N}_{1,\mathcal{E}_i}\}$. Also, we say that a collection $\{\mathcal{N}_{k,l}\}_{l \in [T]}$ for any $k \in [P] \setminus \{1\}$ is *good* if given $k-1$ good collections $\{\mathcal{N}_{h,l}\}_{h \in [k-1], l \in [T]}$, all message symbols $\{X_{i,j} : i \in \mathcal{I}_k, j \in [n]\}$ can be recovered given the

coded symbols $\{Y_{i,m} : i \in [K] \setminus \cup_{g=0}^{h-1} \mathcal{I}_g, m \in \mathcal{N}_{h,\mathcal{E}_i}\}_{h \in [k]}$.

First, we consider the case of $k = 1$, and prove the existence of a good collection $\{\mathcal{N}_{1,l}\}_{l \in [T]}$. For any arbitrary collection $\{\mathcal{N}_{1,l}\}_{l \in [T]}$, it is easy to see that the coded symbols $\{Y_{i,m} : i \in [K], m \in \mathcal{N}_{1,\mathcal{F}_i}\}$ can be recovered given the coded symbols $\{Y_{i,m} : i \in [K], m \in \mathcal{N}_{1,\mathcal{E}_i}\}$. By a standard random linear network coding argument [30], it can be shown that for sufficiently large n (depending on K and q), for randomly chosen $\{\mathcal{N}_{1,l}\}_{l \in [T]}$, the coded symbols $\{Y_{i,m} : i \in [K], m \in \mathcal{N}_{1,\mathcal{F}_i} \cup \mathcal{N}_{1,\mathcal{E}_i}\}$ are linearly independent combinations of the message symbols $\{X_{i,j}\}$ with a nonzero probability. This implies that there exists a collection $\{\mathcal{N}_{1,l}\}_{l \in [T]}$ such that the coded symbols $\{Y_{i,m} : i \in [K], m \in \mathcal{N}_{1,\mathcal{F}_i} \cup \mathcal{N}_{1,\mathcal{E}_i}\}$ are linearly independent combinations of the message symbols $\{X_{i,j}\}$. Fix such a collection $\{\mathcal{N}_{1,l}\}_{l \in [T]}$. Note that $|\{Y_{i,m} : m \in \mathcal{N}_{1,\mathcal{F}_i} \cup \mathcal{N}_{1,\mathcal{E}_i}\}| = |\mathcal{N}_{1,\mathcal{F}_i} \cup \mathcal{N}_{1,\mathcal{E}_i}| = n$ for all $i \in \mathcal{I}_1$. This is because for each $i \in \mathcal{I}_1$, we have (i) $|\mathcal{N}_{1,\mathcal{F}_i}| = \sum_{l \in \mathcal{F}_i} |\mathcal{N}_{1,l}| = \sum_{l \in \mathcal{F}_i} r_1 p_l = r_1 \alpha$; (ii) $|\mathcal{N}_{1,\mathcal{E}_i}| = \sum_{l \in \mathcal{E}_i} |\mathcal{N}_{1,l}| = \sum_{l \in \mathcal{E}_i} r_1 p_l = r_1 \gamma_1$; (iii) $\mathcal{N}_{1,\mathcal{F}_i}$ and $\mathcal{N}_{1,\mathcal{E}_i}$ are disjoint since $\{Y_{i,m} : m \in \mathcal{N}_{1,\mathcal{F}_i} \cup \mathcal{N}_{1,\mathcal{E}_i}\}$ are linearly independent combinations of $\{X_{i,j} : j \in [n]\}$; and (iv) $r_1 \alpha + r_1 \gamma_1 = n$ by the definition of r_1. Note that $\{Y_{i,m} : m \in \mathcal{N}_{1,\mathcal{F}_i} \cup \mathcal{N}_{1,\mathcal{E}_i}\}$ are n linearly independent combinations of the message symbols $\{X_{i,j} : j \in [n]\}$ for each $i \in \mathcal{I}_1$, and $\{Y_{i,m} : m \in \mathcal{N}_{1,\mathcal{F}_i}\}$ can be recovered given $\{Y_{i,m} : m \in \mathcal{N}_{1,\mathcal{E}_i}\}$ for any $i \in \mathcal{I}_1$. This implies that all message symbols $\{X_{i,j} : i \in \mathcal{I}_1, j \in [n]\}$ can be recovered given $\{Y_{i,m} : i \in [K], m \in \mathcal{N}_{1,\mathcal{E}_i}\}$. Thus, $\{\mathcal{N}_{1,l}\}_{l \in [T]}$ is a good collection. Note also that $|\{Y_{i,m} : i \in [K], m \in \mathcal{N}_{1,\mathcal{E}_i}\}| = \sum_{i \in [K]} |\mathcal{N}_{1,\mathcal{E}_i}| = \sum_{i \in [K]} \sum_{l \in \mathcal{E}_i} r_1 p_l = \sum_{i \in [K]} r_1 \beta_i$.

Next, we show that for any $k \in [P] \setminus \{1\}$, given any $k-1$ good collections $\{\mathcal{N}_{h,l}\}_{h \in [k-1], l \in [T]}$, there exists a good collection $\{\mathcal{N}_{k,l}\}_{l \in [T]}$. Consider the case of $k = 2$. Fix a good collection $\{\mathcal{N}_{1,l}\}_{l \in [T]}$. For any arbitrary collection $\{\mathcal{N}_{2,l}\}_{l \in [T]}$, the coded symbols $\{Y_{i,m} : i \in [K], m \in \mathcal{N}_{2,\mathcal{F}_i}\}$ can be recovered given the coded symbols $\{Y_{i,m} : i \in [K], m \in \mathcal{N}_{2,\mathcal{E}_i}\}$. Recall that $\{X_{i,j} : i \in \mathcal{I}_1, j \in [n]\}$ can be recovered given $\{Y_{i,m} : i \in [K], m \in \mathcal{N}_{1,\mathcal{E}_i}\}$. This implies that the coded symbols $\{Y_{i,m} : i \in \mathcal{I}_1, m \in \mathcal{N}_{2,\mathcal{E}_i}\}$ can be recovered given the coded symbols $\{Y_{i,m} : i \in [K], m \in \mathcal{N}_{1,\mathcal{E}_i}\}$. By combining these arguments, one can observe that the coded symbols $\{Y_{i,m} : i \in [K], m \in \mathcal{N}_{1,\mathcal{F}_i}\}$ and the coded symbols $\{Y_{i,m} : i \in [K], m \in \mathcal{N}_{2,\mathcal{F}_i}\}$ can be recovered given the coded symbols $\{Y_{i,m} : i \in [K], m \in \mathcal{N}_{1,\mathcal{E}_i}\}$ and the coded symbols $\{Y_{i,m} : i \in [K] \setminus \mathcal{I}_1, m \in \mathcal{N}_{2,\mathcal{E}_i}\}$. Similarly as before, it can be shown that given a good collection $\{\mathcal{N}_{1,l}\}_{l \in [T]}$, there exists a collection $\{\mathcal{N}_{2,l}\}_{l \in [T]}$ such that the coded symbols $\{Y_{i,m} : i \in [K], m \in \mathcal{N}_{1,\mathcal{F}_i}\}$, $\{Y_{i,m} : i \in [K], m \in \mathcal{N}_{1,\mathcal{E}_i}\}$, $\{Y_{i,m} : i \in [K], m \in \mathcal{N}_{2,\mathcal{F}_i}\}$, and $\{Y_{i,m} : i \in [K] \setminus \mathcal{I}_1, m \in \mathcal{N}_{2,\mathcal{E}_i}\}$ are linearly independent combinations of the message symbols $\{X_{i,j}\}$. Fix such a collection $\{\mathcal{N}_{2,l}\}_{l \in [T]}$. By the same arguments as before, $|\{Y_{i,m} : m \in \mathcal{N}_{1,\mathcal{F}_i} \cup \mathcal{N}_{1,\mathcal{E}_i} \cup \mathcal{N}_{2,\mathcal{F}_i} \cup \mathcal{N}_{2,\mathcal{E}_i}\}| = |\mathcal{N}_{1,\mathcal{F}_i}| + |\mathcal{N}_{1,\mathcal{E}_i}| + |\mathcal{N}_{2,\mathcal{F}_i}| + |\mathcal{N}_{2,\mathcal{E}_i}| = (r_1 + r_2)(\gamma_2 + \alpha) = n$ for all $i \in \mathcal{I}_2$. It is easy to verify that for each $i \in \mathcal{I}_2$, the

coded symbols $\{Y_{i,m} : m \in \mathcal{N}_{1,\mathcal{F}_i} \cup \mathcal{N}_{1,\mathcal{E}_i} \cup \mathcal{N}_{2,\mathcal{F}_i} \cup \mathcal{N}_{2,\mathcal{E}_i}\}$ are n linearly independent combinations of the message symbols $\{X_{i,j} : j \in [n]\}$. Moreover, as we showed earlier, the coded symbols $\{Y_{i,m} : i \in [K], m \in \mathcal{N}_{1,\mathcal{F}_i}\}$ and $\{Y_{i,m} : i \in [K], m \in \mathcal{N}_{2,\mathcal{F}_i}\}$ can be recovered given the coded symbols $\{Y_{i,m} : i \in [K], m \in \mathcal{N}_{1,\mathcal{E}_i}\}$ and $\{Y_{i,m} : i \in [K] \setminus \mathcal{I}_1, m \in \mathcal{N}_{2,\mathcal{E}_i}\}$. Putting these arguments together, it follows that all message symbols $\{X_{i,j} : i \in \mathcal{I}_2, j \in [n]\}$ can be recovered given the coded symbols $\{Y_{i,m} : i \in [K] \setminus \cup_{g=0}^{h-1}\mathcal{I}_g, m \in \mathcal{N}_{h,\mathcal{E}_i}\}_{h\in[2]}$. This implies that $\{\mathcal{N}_{2,l}\}_{l\in[T]}$ is a good collection. Note also that $|\{Y_{i,m} : i \in [K] \setminus \cup_{g=0}^{h-1}\mathcal{I}_g, m \in \mathcal{N}_{h,\mathcal{E}_i}\}_{h\in[2]}| = \sum_{i\in[K]}|\mathcal{N}_{1,\mathcal{E}_i}| + \sum_{i\in[K]\setminus\mathcal{I}_1}|\mathcal{N}_{2,\mathcal{E}_i}| = \sum_{i\in[K]}r_1\beta_i + \sum_{i\in[K]\setminus\mathcal{I}_1}r_2\beta_i = \sum_{i\in\mathcal{I}_1}r_1\beta_i + \sum_{i\in[K]\setminus\mathcal{I}_1}(r_1+r_2)\beta_i$.

Repeating the same arguments as above for the cases of $k = 3, \ldots, P$, it can be shown that for each $k \in [P] \setminus \{1\}$, given any $k-1$ good collections $\{\mathcal{N}_{h,l}\}_{h\in[k-1],l\in[T]}$, there exists a good collection $\{\mathcal{N}_{k,l}\}_{l\in[T]}$. That is, there exist $\{\mathcal{N}_{k,l}\}_{l\in[T]}$ such that all message symbols $\{X_{i,j} : i \in \mathcal{I}_k, j \in [n]\}$ can be recovered given the coded symbols $\{Y_{i,m} : i \in [K] \setminus \cup_{g=0}^{h-1}\mathcal{I}_g, m \in \mathcal{N}_{h,\mathcal{E}_i}\}_{h\in[k]}$. This implies that for each $k \in [P]$, there exist $\{\mathcal{N}_{h,l}\}_{h\in[k],l\in[T]}$ such that all message symbols $\{X_{i,j} : i \in \cup_{h=1}^{k}\mathcal{I}_h, j \in [n]\}$ can be recovered given the coded symbols $\{Y_{i,m} : i \in [K] \setminus \cup_{g=0}^{h-1}\mathcal{I}_g, m \in \mathcal{N}_{h,\mathcal{E}_i}\}_{h\in[k]}$. Using the same arguments as before, it can be shown that for each $k \in [P]$, $|\{Y_{i,m} : i \in [K] \setminus \cup_{g=0}^{h-1}\mathcal{I}_g, m \in \mathcal{N}_{h,\mathcal{E}_i}\}_{h\in[k]}| = \sum_{h\in[k]}\sum_{i\in[K]\setminus(\mathcal{I}_0\cup\ldots\cup\mathcal{I}_{h-1})}r_h\beta_i$. Taking $k = P$, it then follows that all Kn message symbols $\{X_{i,j}\}$ can be recovered given the $\sum_{h\in[P]}\sum_{i\in[K]\setminus(\mathcal{I}_0\cup\ldots\cup\mathcal{I}_{h-1})}r_h\beta_i$ coded symbols $\{Y_{i,m} : i \in [K] \setminus \cup_{g=0}^{h-1}\mathcal{I}_g, m \in \mathcal{N}_{h,\mathcal{E}_i}\}_{h\in[P]}$. It is also easy to verify that $\sum_{h\in[P]}\sum_{i\in[K]\setminus(\mathcal{I}_0\cup\ldots\cup\mathcal{I}_{h-1})}r_h\beta_i = \sum_{h\in[P]}\sum_{i\in\mathcal{I}_h}(r_1+\cdots+r_h)\beta_i = (\sum_{i\in[K]}(\beta_i/(\alpha+\beta_i)))n$. By combining these arguments, it follows that there exist $(\sum_{i\in[K]}(\beta_i/(\alpha+\beta_i)))n$ coded symbols given which all Kn message symbols can be recovered.

To complete the proof, we need to show that $(\sum_{i\in[K]}(\beta_i/(\alpha+\beta_i)))n \le (M/(D+M))kn$. It is easy to show that $\sum_{i\in[K]}(\beta_i/(\alpha+\beta_i))$ is maximized when $\beta_i = M/K$ for all $i \in [K]$, noting that $\alpha = D/K$ does not depend on $\{\beta_i\}_{i\in[K]}$, and $\sum_{i\in[K]}\beta_i = M$. This readily implies that $\sum_{i\in[K]}(\beta_i/(\alpha+\beta_i)) \le (M/(D+M))K$.

V. Proof of Theorem 2

In this section, we present a linear IPIR protocol, referred to as *Group-and-Code*, for all K, D, M such that $(D/R + M/R) \mid K$, where $R := \gcd(D, M)$, and show that this protocol achieves the rate $(D+M)/K$. The Group-and-Code protocol consists of three steps described below.

For ease of notation, we define $d := D/R$, $m := M/R$, $T := d + m$, and $P := K/T$.

Step 1: First, the user randomly partitions the message indices $1, \ldots, K$ into P groups $\mathcal{I}_1, \ldots, \mathcal{I}_P$, each of size T as follows: (i) R (out of P) groups are chosen at random, and each of these groups is filled with d randomly chosen message indices from W and m randomly chosen message

indices from S; and (ii) the remaining $P - R$ groups are randomly filled with the remaining $K - D - M$ message indices from $[K] \setminus (W \cup S)$. Next, the user constructs d arbitrary length-T row-vectors $v_1, \ldots, v_d$ with entries from $\mathbb{F}_q$ such that the matrix $V := [v_1^\mathsf{T}, \ldots, v_d^\mathsf{T}]^\mathsf{T}$ generates a $[T, d]$ MDS code over $\mathbb{F}_q$. The user then sends $\mathcal{I}_1, \ldots, \mathcal{I}_P$ and $v_1, \ldots, v_d$ as the query $\mathrm{Q}^{[W,S]}$ to the server.

Step 2: Given the query $\mathrm{Q}^{[W,S]}$, for each $k \in [P]$ and each $l \in [d]$, the server computes $Z_{k,l} = \sum_{j\in[T]}v_{l,j}X_{i_{k,j}}$, where $\mathcal{I}_k = \{i_{k,1}, \ldots, i_{k,T}\}$ and $v_l = [v_{l,1}, \ldots, v_{l,T}]$, and sends $\{Z_{k,l}\}_{k\in[P],l\in[d]}$ back to the user as the answer $\mathrm{A}^{[W,S]}$.

Step 3: Given the answer $\mathrm{A}^{[W,S]}$ and the side information X_S, the user recovers their demand X_W as follows. Without loss of generality, assume that $\mathcal{I}_1, \ldots, \mathcal{I}_R$ contain messages indices from $W \cup S$. For each $k \in [R]$, let $W_k \subset \mathcal{I}_k$ and $S_k \subset \mathcal{I}_k$ be such that (i) $W_k \subseteq W$ and $S_k \subseteq S$, and (ii) $\mathcal{I}_k = W_k \cup S_k$. For each $k \in [R]$ and each $l \in [d]$, the user computes $\tilde{Z}_{k,l}$ by subtracting off the contribution of the m side information messages X_{S_k} from $Z_{k,l}$, i.e., $\tilde{Z}_{k,l} = \sum_{j\in\mathcal{J}_k}v_{l,j}X_{i_{k,j}}$, where $\mathcal{J}_k \subset [T]$ is such that $W_k = \{i_{k,j} : j \in \mathcal{J}_k\}$. For each $k \in [R]$, the user then recovers the d demand messages X_{W_k} from $\tilde{Z}_{k,1}, \ldots, \tilde{Z}_{k,d}$ by solving a system of linear equations.

The rate of the Group-and-Code protocol is equal to $(D+M)/K$. Note that for any realization (W,S), the answer $\mathrm{A}^{[W,S]}$ consists of $L := Pd$ $(= Kd/(d+m) = KD/(D+M))$ linear combinations $\{Z_{k,l}\}_{k\in[P],l\in[d]}$. It is also easy to see that $\{Z_{k,l}\}_{k\in[P],l\in[d]}$ are linearly independent combinations of the messages. This implies that all realizations of $\mathbf{A}^{[\mathbf{W},\mathbf{S}]}$ consist of L linearly independent combinations of the messages. Thus, $H(\mathbf{A}^{[\mathbf{W},\mathbf{S}]}) = LB$ (as discussed in Section II), where $B = H(\mathbf{X}_i)$ for all $i \in [K]$. Since $H(\mathbf{X}_\mathbf{W}) = DB$, then the rate of this protocol is equal to $DB/(LB) = D/L = (D+M)/K$.

It is also easy to see that the recoverability condition is satisfied. Fix an arbitrary $k \in [R]$. Note that $\tilde{Z}_{k,1}, \ldots, \tilde{Z}_{k,d}$ are linear combinations of the demand messages X_{W_k}. Note also that the coefficient vectors of these linear combinations are the rows of a $d \times d$ submatrix of V, and every $d \times d$ submatrix of V is full-rank because V generates a $[T, d]$ MDS code. Thus, $\tilde{Z}_{k,1}, \ldots, \tilde{Z}_{k,d}$ are d linearly independent combinations of the d demand messages X_{W_k}.

To prove that the individual privacy condition is satisfied, we need to show that $\mathbb{P}(i \in \mathbf{W}|\mathbf{Q}^{[\mathbf{W},\mathbf{S}]} = \mathrm{Q}^{[W,S]}) = D/K$ for all $i \in [K]$. Fix an arbitrary $i \in [K]$. Let $k \in [P]$ be such that $i \in \mathcal{I}_k$. From the description of the protocol, it is easy to see that $\mathbb{P}(i \in \mathbf{W}|\mathbf{Q}^{[\mathbf{W},\mathbf{S}]} = \mathrm{Q}^{[W,S]})$ is equal to the probability that the kth group is one of the R groups that are randomly chosen at first and the message index i is one of the d demand message indices that are placed in the kth group (i.e., the message index i belongs to W_k). The probability that the kth group is one of the R chosen groups is $R/P = (D+M)/K$, and given that the kth group is one of the R chosen groups, the probability that the message index i belongs to W_k is $d/T = D/(D+M)$. Thus, $\mathbb{P}(i \in \mathbf{W}|\mathbf{Q}^{[\mathbf{W},\mathbf{S}]} = \mathrm{Q}^{[W,S]}) = (R/P) \times (d/T) = D/K$.

REFERENCES

[1] A. Heidarzadeh, S. Kadhe, S. E. Rouayheb, and A. Sprintson, "Single-server multi-message individually-private information retrieval with side information," in *2019 IEEE International Symposium on Information Theory (ISIT)*, July 2019, pp. 1042–1046.

[2] A. Heidarzadeh and A. Sprintson, "Single-server individually-private information retrieval: A combinatorial approach," in *2021 IEEE Information Theory Workshop (ITW)*, 2021, pp. 1–6.

[3] K. Banawan and S. Ulukus, "Multi-message private information retrieval," in *2017 IEEE International Symposium on Information Theory (ISIT)*, June 2017, pp. 1898–1902.

[4] K. Banawan and S. Ulukus, "Multi-message private information retrieval: Capacity results and near-optimal schemes," *IEEE Transactions on Information Theory*, vol. 64, no. 10, pp. 6842–6862, Oct 2018.

[5] M. H. Mousavi, M. Ali Maddah-Ali, and M. Mirmohseni, "Private inner product retrieval for distributed machine learning," in *2019 IEEE International Symposium on Information Theory (ISIT)*, 2019, pp. 355–359.

[6] A. Heidarzadeh, N. Esmati, and A. Sprintson, "Single-server private linear transformation: The joint privacy case," *IEEE Journal on Selected Areas in Communications*, pp. 1–1, 2022.

[7] F. Kazemi, E. Karimi, A. Heidarzadeh, and A. Sprintson, "Single-server single-message online private information retrieval with side information," in *2019 IEEE International Symposium on Information Theory (ISIT)*, July 2019, pp. 350–354.

[8] A. Heidarzadeh, N. Esmati, and A. Sprintson, "Single-server private linear transformation: The individual privacy case," June 2021. [Online]. Available: arXiv:2106.05222

[9] H. Sun and S. A. Jafar, "The capacity of private information retrieval," *IEEE Transactions on Information Theory*, vol. 63, no. 7, pp. 4075–4088, July 2017.

[10] R. Tandon, "The capacity of cache aided private information retrieval," in *55th Annual Allerton Conf. on Commun., Control, and Computing*, Oct 2017, pp. 1078–1082.

[11] R. Tajeddine and S. El Rouayheb, "Robust private information retrieval on coded data," in *2017 IEEE International Symposium on Information Theory (ISIT)*, 2017.

[12] R. Tajeddine, O. W. Gnilke, D. Karpuk, R. Freij-Hollanti, C. Hollanti, and S. E. Rouayheb, "Private information retrieval schemes for coded data with arbitrary collusion patterns," in *2017 IEEE International Symposium on Information Theory (ISIT)*, June 2017, pp. 1908–1912.

[13] H. Sun and S. A. Jafar, "The capacity of robust private information retrieval with colluding databases," *IEEE Transactions on Information Theory*, vol. 64, no. 4, pp. 2361–2370, April 2018.

[14] Y. Wei, K. Banawan, and S. Ulukus, "Cache-aided private information retrieval with partially known uncoded prefetching: Fundamental limits," *IEEE Journal on Selected Areas in Communications*, vol. 36, no. 6, pp. 1126–1139, June 2018.

[15] ——, "Fundamental limits of cache-aided private information retrieval with unknown and uncoded prefetching," *IEEE Transactions on Information Theory*, pp. 1–1, 2018.

[16] C. Tian, H. Sun, and J. Chen, "Capacity-achieving private information retrieval codes with optimal message size and upload cost," *IEEE Transactions on Information Theory*, vol. 65, no. 11, pp. 7613–7627, 2019.

[17] S. Kadhe, B. Garcia, A. Heidarzadeh, S. El Rouayheb, and A. Sprintson, "Private information retrieval with side information," *IEEE Transactions on Information Theory*, vol. 66, no. 4, pp. 2032–2043, 2020.

[18] A. Heidarzadeh, F. Kazemi, and A. Sprintson, "Capacity of single-server single-message private information retrieval with coded side information," in *2018 IEEE Information Theory Workshop (ITW)*, Nov 2018, pp. 1–5.

[19] ——, "Capacity of single-server single-message private information retrieval with private coded side information," in *2019 IEEE International Symposium on Information Theory (ISIT)*, July 2019, pp. 1662–1666.

[20] ——, "The role of coded side information in single-server private information retrieval," *IEEE Transactions on Information Theory*, vol. 67, no. 1, pp. 25–44, 2021.

[21] F. Kazemi, E. Karimi, A. Heidarzadeh, and A. Sprintson, "Multi-server private information retrieval with coded side information," in *2019 16th Canadian Workshop on Information Theory (CWIT)*, 2019, pp. 1–6.

[22] ——, "Private information retrieval with private coded side information: The multi-server case," in *2019 57th Annual Allerton Conference on Communication, Control, and Computing (Allerton)*, 2019, pp. 1098–1104.

[23] Z. Chen, Z. Wang, and S. A. Jafar, "The capacity of T-private information retrieval with private side information," *IEEE Transactions on Information Theory*, vol. 66, no. 8, pp. 4761–4773, 2020.

[24] S. Li and M. Gastpar, "Converse for multi-server single-message pir with side information," in *2020 54th Annual Conference on Information Sciences and Systems (CISS)*, 2020, pp. 1–6.

[25] M. Shrivastava and P. Sarvepalli, "Capacity Achieving Uncoded PIR Protocol based on Combinatorial Designs," *arXiv e-prints*, p. arXiv:2103.09804, Mar. 2021.

[26] M. K. K. H. and J. Harshan, "XOR-Based Codes for Private Information Retrieval with Private Side Information," *arXiv e-prints*, p. arXiv:2105.05788, May 2021.

[27] A. Heidarzadeh, S. Kadhe, B. Garcia, S. E. Rouayheb, and A. Sprintson, "On the capacity of single-server multi-message private information retrieval with side information," in *2018 56th Annual Allerton Conf. on Commun., Control, and Computing*, Oct 2018.

[28] S. P. Shariatpanahi, M. J. Siavoshani, and M. A. Maddah-Ali, "Multi-message private information retrieval with private side information," in *2018 IEEE Information Theory Workshop (ITW)*, 2018, pp. 1–5.

[29] S. Li and M. Gastpar, "Single-server multi-message private information retrieval with side information," in *2018 56th Annual Allerton Conf. on Commun., Control, and Computing*, Oct 2018.

[30] T. Ho, M. Medard, R. Koetter, D. Karger, M. Effros, J. Shi, and B. Leong, "A random linear network coding approach to multicast," *IEEE Transactions on Information Theory*, vol. 52, no. 10, pp. 4413–4430, 2006.